

Lexi™ Topic Models

Setting up Topic Models in the EEG Cloud

Using Topic Models helps Lexi to learn about specific names, places, and topics unique to your application. This document will explain how Topic Models work and how to maximize Lexi's performance by creating effective Topic Models.

Getting Started

All Topic Models are controlled and maintained remotely through your EEG Cloud account login at eegcloud.tv – credentials are provided by EEG Support with Lexi activated encoder purchases and demos. Once logged in, **navigate to the Lexi Automatic Live Captioning app from your home dashboard.**

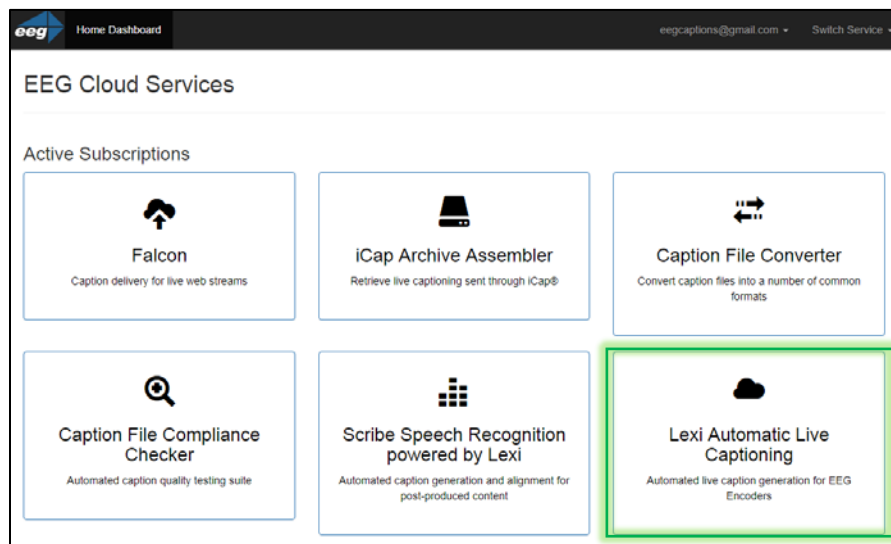


Figure 1: EEG Cloud Services Dashboard

Once inside the Lexi Application **select “Topic Models” from the top menu** to access the Topic Model Manager.

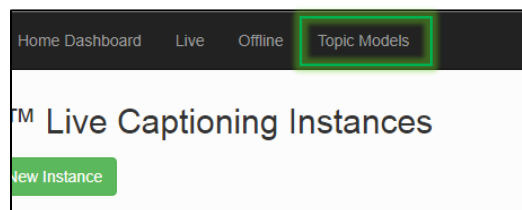


Figure 2: Topic Models Button Located on Top Menu

Adding an EEG Topic Model

1. Currently EEG has two pre-made Topic Models available – one for Legislative applications and another for US News applications. Models for other applications will become available in the future. **If your topic matter is not represented in the list navigate to the [Creating Your Own Topic Model](#) section of this document.** From the Topic Model Manager homepage click the green “Add a Topic Model” button, then select “Choose one of our Topic Models”.
2. Select the EEG Model you wish to add to your account from the dropdown, enter a name for the model (in this example EEG News), and then click “Create Model”.

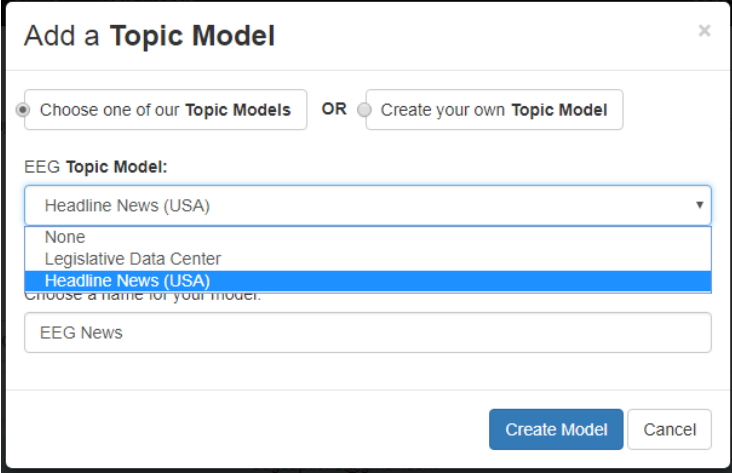


Figure 2: Selecting an EEG Topic Model

3. While the model is being created you will be taken to its management page (depicted in Figure 3) where the model can be further customized if desired. **It will take a few moments for the model to finish creating before you can click anything or make changes to it. A blue indicator with a spinning gear will be present at the top of the page for the duration of the models creation.**

While EEG Topic Models serve as a great base for achieving a high level of caption accuracy in your program, **you may want to dial-in the model further to ensure recognition of specific names, locations, or other vocabulary unique to your application.**

For example: A local news station using EEG’s “US Headline News” topic model as a base may also want add in their own anchor names, town names, and the names of local figures to ensure they are observed by the system. Adding these names or any combination of teleprompter data, scripts, or relevant websites will enable you to do this.

To add individual words to the system click on the “Vocab Control” tab and add words to your vocabulary list. To have Lexi analyze a website or a text file, click either the “Learn from a Website” or “Learn from a File” button and follow the prompts. **More detailed instructions on adding websites or files and using vocab control can be found in the [Topic Model Training / Best Practices](#) and [Vocab Control / Phonetic Training](#) sections of this document.**

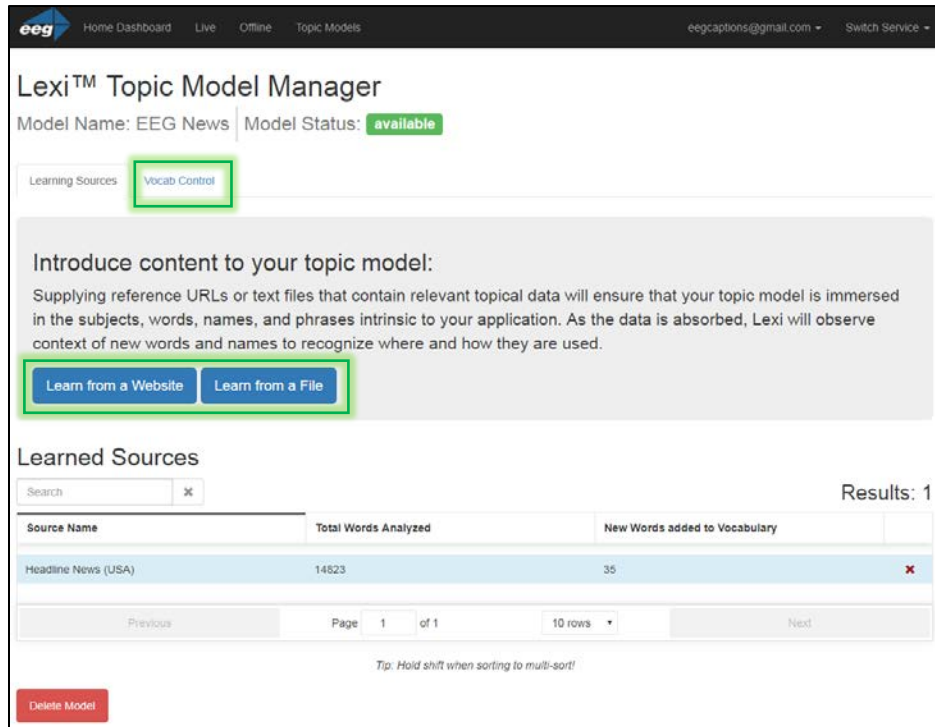


Figure 3: Customizing Your EEG Topic Model

Creating Your Own Topic Model

1. From the Topic Model Manager homepage click the green “Add a Topic Model” button, then select “Create your own Topic Model”. Give your model a topical or memorable name (*Basketball* in our example) and click “Create Model” to continue.

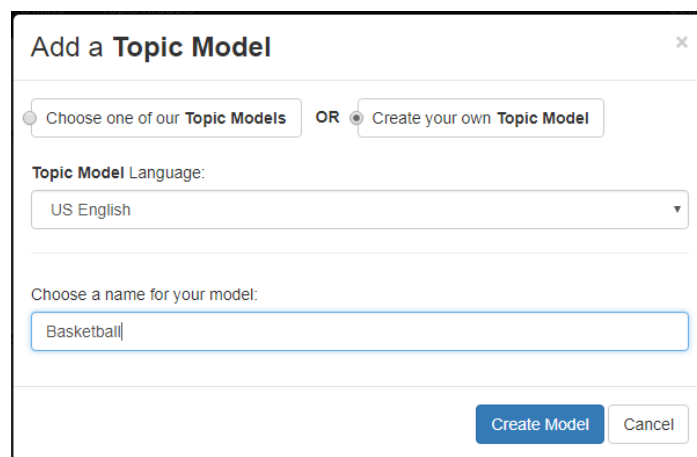


Figure 4: Creating your own Topic Model

- Once the shell of your model has been created you will be taken to its management page where you will train it using any combination of website data, text files, or by adding individual words and names. To add individual words to the system click on the “Vocab Control” tab and add words to your vocabulary list. To have Lexi analyze a website or a text file, click either the “Learn from a Website” or “Learn from a File” button and follow the prompts. **More detailed instructions on adding websites or files and using vocab control can be found in the [Topic Model Training / Best Practices](#) and [Vocab Control / Phonetic Training](#) sections of this document.**

Topic Model Training / Best Practices

Lexi loves to read – so the more relevant and contextual information supplied in your Topic Model the better. For example: a list of names, while somewhat helpful, can often be less effective than providing an article or content where these names are used in context. **Introducing content to your Topic Model with a Website (or more than 1) to scan and absorb information from is often the most effective method** for training Lexi. The other option offered is to learn from a text file.

Learning Content from a Website

- Add the FULL desired URL** to the URL section as in `http://www.example.com`
- If “All pages in this domain” is selected** Lexi will automatically select up to 50 pages to observe from the path of the provided URL. That being the case, either be sure to provide a strategic URL that will not encourage Lexi to roam a broad range of unrelated paths **OR use the advanced options** to specify which paths to scan and which paths to ignore.
- Advanced Options** enable further control of the scanning process. In the example shown in figure 5, “*espn.com*” itself covers a very wide range of various sports topics not specific to basketball and on its own is too broad to be a useful resource. However, **by using the Whitelist and Blacklist features in Advanced Options** Lexi is instructed to only scan the pages stemming from the “*espn.com/nba*” path and to ignore its subpaths “*espn.com/nba/scoreboard*” and “*espn.com/nba/statistics*” (because these subpaths provide irrelevant numerical data).

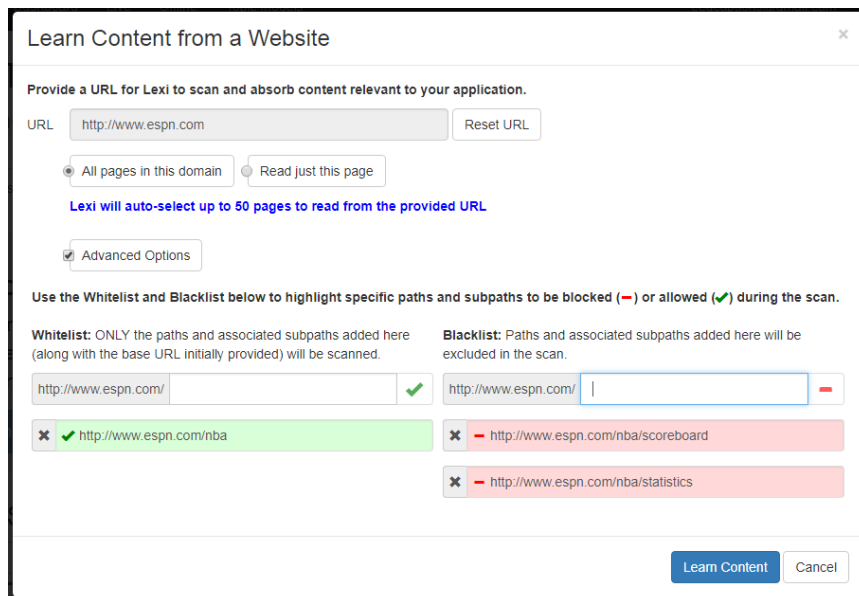


Figure 5: Learn Content from a Website

4. If **“Read just this page”** is selected Lexi will scan only the single page of the provided URL.
5. Click **“Learn Content”** to commence the scan. Lexi will begin analyzing the web content immediately and will display an indicator as pictured in Figure 6 for the duration of the scan. Scans can take several minutes and may depend on the amount of content provided. **Once completed the Source name and the results of the scan will appear in the “Learned Sources” list of your Topic Model.**

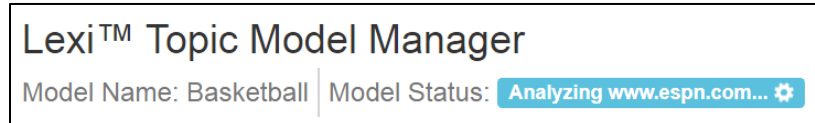


Figure 6: Analyzing Indicator

Learned Sources			Results: 1
Source Name	Total Words Analyzed	New Words added to Vocabulary	
www.espn.com	2030	8	✖

Previous Page 1 of 1 10 rows Next

Figure 7: Learned Sources

Learning Content from a File

1. **Click the blue “Learn from a File” button.** In the resulting popup window click **“Select a File”** and browse for the text file you would like to upload.
2. **For best results, upload files that contain lots of context.** In a broadcast news example; a text file list of anchor names, while useful, will be less effective than providing a script or two of a past airing where these names were used in context.

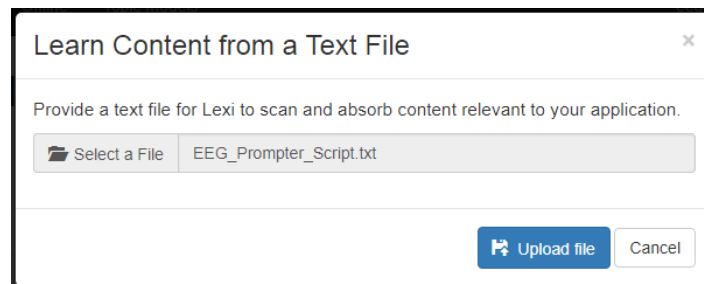
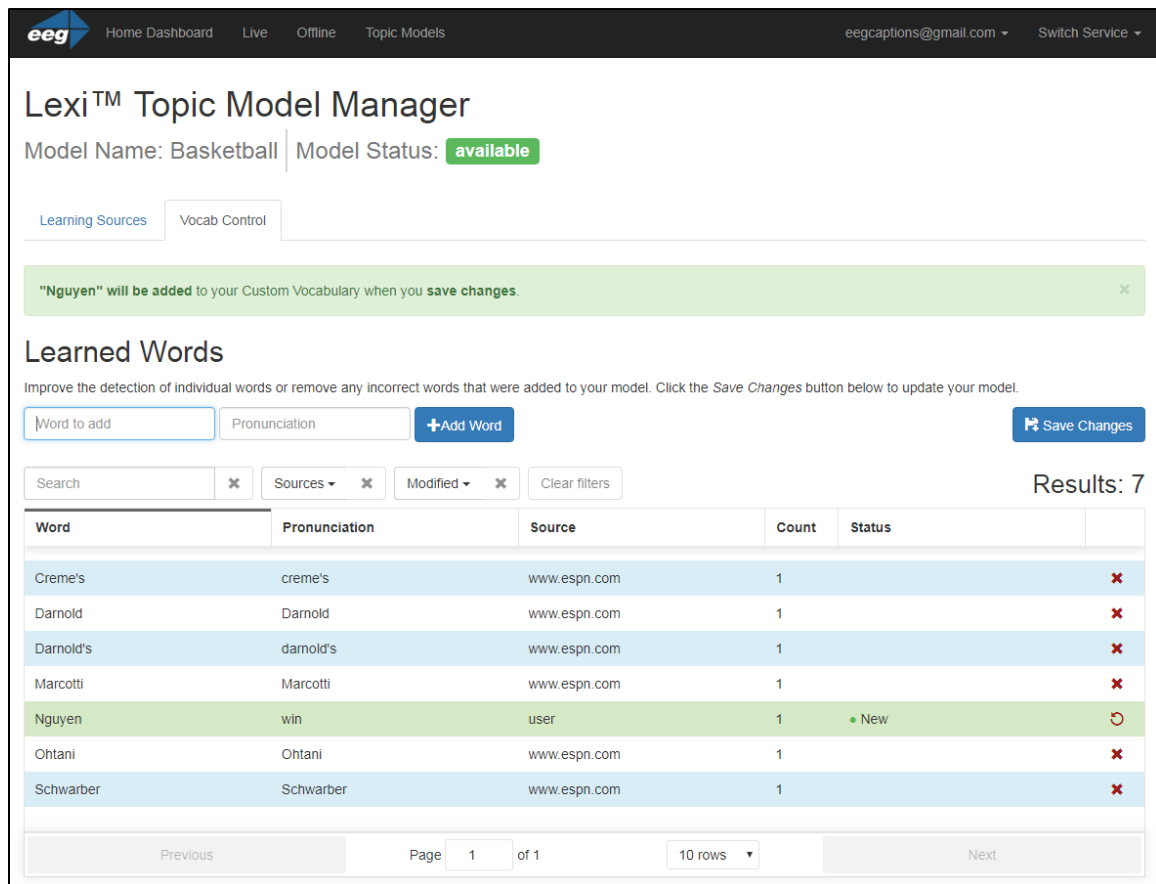


Figure 8: Text File Uploader

Vocab Control / Phonetic Training

The “Vocab Control” tab displays a list of your model’s Learned words along with phonetic spelling for each. Here, you can add new or edit existing words to encourage Lexi’s accuracy further. For instance, if Lexi is not grasping a certain basketball player or news anchor’s uncommon name – editing the phonetic spelling to your specification may help Lexi to get it right in the future. **For example, the last name Nguyen is pronounced as “win” – so adding “win” as the phonetic spelling should assist Lexi in accurately recognizing that name.** Be sure to save all changes after adding / editing words in the Vocab Control section.



Lexi™ Topic Model Manager
 Model Name: Basketball | Model Status: available

Learning Sources | Vocab Control

"Nguyen" will be added to your Custom Vocabulary when you save changes.

Learned Words
 Improve the detection of individual words or remove any incorrect words that were added to your model. Click the Save Changes button below to update your model.

Word to add: Pronunciation: +Add Word Save Changes

Search Sources Modified Clear filters Results: 7

Word	Pronunciation	Source	Count	Status
Creme's	creme's	www.espn.com	1	
Darnold	Darnold	www.espn.com	1	
Darnold's	darnold's	www.espn.com	1	
Marcotti	Marcotti	www.espn.com	1	
Nguyen	win	user	1	New
Ohtani	Ohtani	www.espn.com	1	
Schwarber	Schwarber	www.espn.com	1	

Previous Page 1 of 1 10 rows Next

Figure 9: Adding and Editing Phonetic Spelling using Vocab Control

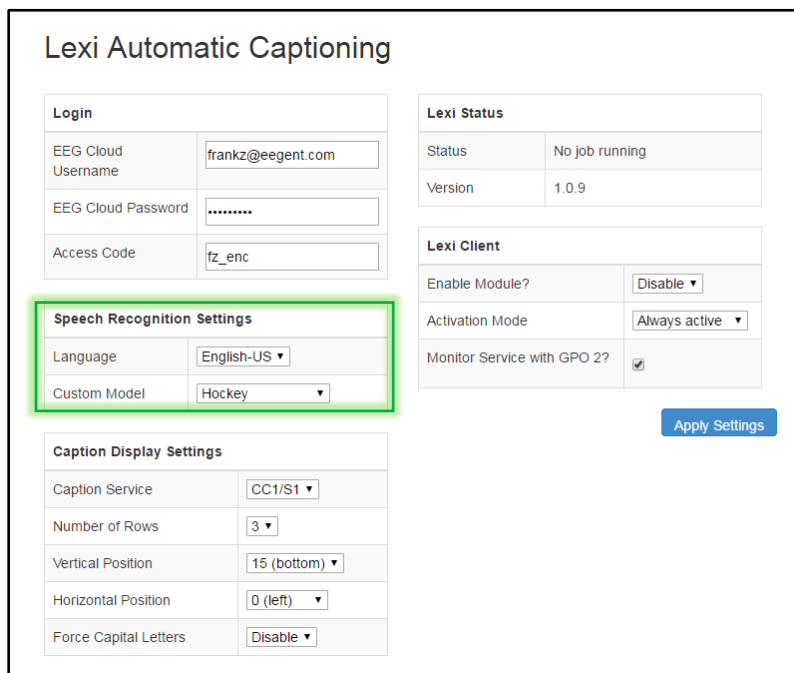
Accessing Your Custom Model From Your EEG Encoder

Once you’ve created and saved your Topic Model in the EEG Cloud – you will need to access it from your EEG encoder interface to experience the results in your live broadcast captioning. Accessing your new speech model from the encoder will require the following steps:

1. **Navigate to the encoder’s web interface.** Accessed by entering the IP address assigned to your encoder in your computers web browser on the same network. If you set the encoder up with a static IP address, the address for your web interface will be the same. If you set the encoder up with DHCP, the address can change at will and you should check the IP from the front panel of your

encoder at **System Setup > Network > IP Address** to ensure you are using the correct IP to access the web interface.

2. **Select Lexi Automatic Captioning** from the list of modules on the left side of the screen.
3. **Ensure your EEG Cloud Username and Password and Access Code are entered** in the Lexi Module interface.
4. **Under Speech Recognition Settings, select your desired custom model** from the Custom Model dropdown list.



Login	
EEG Cloud Username	frankz@eegent.com
EEG Cloud Password
Access Code	fz_enc

Lexi Status	
Status	No job running
Version	1.0.9

Lexi Client	
Enable Module?	Disable ▾
Activation Mode	Always active ▾
Monitor Service with GPO 2?	<input checked="" type="checkbox"/>

Speech Recognition Settings	
Language	English-US ▾
Custom Model	Hockey ▾

Caption Display Settings	
Caption Service	CC1/S1 ▾
Number of Rows	3 ▾
Vertical Position	15 (bottom) ▾
Horizontal Position	0 (left) ▾
Force Capital Letters	Disable ▾

Apply Settings

Figure 10: Accessing Your Topic Model through the EEG Encoder Web Interface